

Title: The scope of language contact as a constraint factor in language change: the periphrasis *haber de* + infinitive in a corpus of language immediacy in modern Spanish.

Author:

José Luis Blas Arroyo (Universitat Jaume I. E-mail: blas@fil.uji.es. Phone number: 964729625).

Short Title: The scope of language contact a constraint factor in language change: the Spanish periphrasis *haber de* + infinitive

The scope of language contact as a constraint factor in language change: the periphrasis *haber de* + infinitive in a corpus of language immediacy in modern Spanish

Abstract:

In this work an empirical study grounded in the principles and methods of the comparative variationist framework is conducted to measure the scope of language contact as a factor constraining some potentially diverging uses of a Spanish verbal periphrasis that has undergone a sharp decline over the last century (*haber de* + infinitive). The analysis is based on three independent samples of text that correspond to three dialectal areas of peninsular Spanish (monolingual zones, Catalan-speaking linguistic territories and the north-western linguistic area). These samples, extracted from a corpus made up of texts of communicative immediacy from the 19th and the first half of the 20th centuries, confirm the existence of a certain linguistic convergence in the expressive habits of the speakers in the bilingual communities. In each region, however, the outcomes are different, due to parallel differences in the structural position of the periphrasis in each language. However, a thorough analysis of the variable context that surrounds the periphrasis shows that the observed differences do not affect the essence of the underlying grammar of this variant, whose decline (which favours *tener que* + infinitive and becomes faster as the 20th century advances) is constrained by identical linguistic and extralinguistic conditioning factors in all the dialectal areas.

Key words: modal periphrases, language contact, historical sociolinguistics, language change, corpus linguistics, texts of communicative immediacy, Spanish, social factors.

1. INTRODUCTION

In several previous works we have seen how the old alternation between the Spanish periphrases *haber de* and *tener que* + infinitive, exemplified in (1) to (4), have been acting within the same sphere of modality for centuries in a process of language change that, following a clear predominance of the former throughout most of the history of the Spanish language, has finally quite clearly opted for the latter (Author and Author X, a; Author and Author X, b):

(1) ... el portador de esta le dirá a usted lo que hay en el particular, y ese le dará el norte onde para ese caballero, y le enterará de todo lo que *ha de hazer*, y así enteniéndolos yo acá sabré lo que *e de hazer* (1817; *La emigración en tinta y papel*)¹

[...the bearer of this will tell you what there is in particular, and he will tell you where this gentleman is, and he will tell you what must be done, and thus by understanding them here I shall know what I have to do]

(2) A Blas no *tiene que acerle* ninguna rropa porque aquí de nada sirbe (año 1816; *La emigración en tinta y papel*)

[You don't have to make any clothes for Blas because here they are of no use at all]

(3) Y yo al oír estas oraciones encima de lo mío me hacía llorar [...] llorar porque seré incapaz de pagar las obligaciones devidas á mis queridos papás, pero *he de ir haciendo* medios poco a poco para siquiera no darles disgustos... (1950; *Cartas desde América*)

[And hearing these prayers on top of mine made me cry (...) cry because I will be unable to pay back what I owe my dear parents, but I have to gradually take steps to ensure I don't upset them...]

(4) Ahora se me quiere enviar a Valencia para dedicarme a lo mismo, pero mi renuncia de lo de aquí *tendría que hacerla* allí de no plantearse la propaganda como es debido, por lo que para ir a Valencia exijo condiciones mínimas de eficacia (1946; *Dramas de refugiados*).

[Now they want to send me to Valencia to spend my time doing the same thing, but my renunciation of what goes on here would have to be done there if the propaganda is not as it should be; therefore to go to Valencia I demand the minimum conditions of effectiveness]

From a corpus made up of texts of communicative immediacy,² we have seen how the uses of *haber de* in the mid-20th century underwent a very considerable decline with respect to those of previous centuries. This decline is particularly pronounced in relation to classical Spanish, but also even in comparison to the 19th century, when the periphrasis still retained a considerable amount of vitality. Nevertheless, this loss of prominence of what had been the dominant periphrasis since the Middle Ages occurs in practically all linguistic contexts, the result being that it was replaced by its competitor, *tener que* – and to a lesser extent, *deber (de)* – as the prevalent form in nearly all of them. Despite the waning of *haber de*, a careful analysis of the variable context reveals the existence of certain points in the grammar that still favour it and, what is more interesting, the fact that, with the odd exception, these points are practically identical to those that operated in the past. Hence, factors like the length of the verb group, the verb tense and mood, and manner of action of the main verb or the morphemes indicating person and number still condition the selection of the periphrasis in the same way they did in the past. Nonetheless, these are not the only structural factors involved in this variation phenomenon. On the stylistic axis we have also seen how the distinction

¹ The examples are reproduced here with the spelling used in the original texts.

² This work is included within the Research Project “xxxxxxxxxxxxxxxxxxxxxxxxxxxx”, funded by the Spanish Ministry of Science and Technology, which is being carried out by a research team led by the author (period 2011-2013). For a review of the main features and aims of this historical sociolinguistics project, see Author (forthcoming).

between different types of letters (private/distant letters) and autobiographical texts (memoirs, diaries, etc.) exerts a decisive influence in the last two centuries. Furthermore, these results anticipate the situation in the present-day language in which *haber de* is mostly limited to some formal contexts of the written language, something that has been addressed in the literature on linguistics on a number of occasions (Seco, 1986; Gómez Torrego, 1988, 1999; Hernández García, 1998; Fernández de Castro, 1999; Sinner, 2003; Martínez Díaz, 2002, 2003, 2008; García Fernández, 2006, 2012; Hernández Díaz, 2006; López Izquierdo, 2008).

Although less frequently, that same literature also contains references to notable contrasts in the realisation of the periphrasis from the diatopic point of view. In this regard, for example, attention has been drawn to the potential influence of language contact in this area of the infinitive verbal periphrases in regions such as Catalonia or Galicia, where Spanish has coexisted for centuries alongside another Romance language whose paradigms display certain significant differences compared to those of Castilian Spanish (Álvarez, 1983; Wesch, 1997; Álvarez et al., 1998; Rojo, 1974, 2004; Martínez Díaz, 2002, 2003, 2008; Author, 2004; Sinner, 2003, 2004; Sinner and Wesch, 2008; Hernández García, 1998). Nevertheless, the data about this possible influence are, generally speaking, scarce and fragmented, when not simply limited to subjective impressions. Besides, they are not the result of an exhaustive review of the variable context, in which all the potentially involved factors are considered at the same time.

The objectives of this research focus precisely on the systematic analysis of this variable context, with a view to determining the exact extent to which that secular contact between different related languages could influence the realisation of these infinitive verbal periphrases. By using a corpus of communicative immediacy that includes representative samples of several dialectal areas of peninsular Spanish, in this work we intend to find answers to the following questions: are there any differences in the realisation of the periphrases, and particularly of the receding variant *haber de*, from one region to another in the last two centuries? If the answer is affirmative: what is the magnitude of those differences? Are they some differences of degree or, on the contrary, do they hide within them diverging underlying grammars?

Before going on to answer these questions by analysing the data derived from a study based on the principles of the comparative variationist framework (section 4), in the next section we will review the key points of contact and structural conflict of the periphrases in the different languages involved.³ In section 3, we will outline the main methodological details of the research and conclude with some reflections on the real role of linguistic convergence in this variation phenomenon (section 5).

2. INFINITIVE VERB PERIPHRASES IN THE ROMANCE LANGUAGES OF THE PENINSULA: POINTS OF CONTACT AND POINT OF STRUCTURAL CONFLICT

2.1 An age-old change in progress in the Spanish syntax: the *haber de/tener que* + infinitive alternation

The origins of modal periphrases of obligation lie in the corresponding Latin constructions with the verbs *debeo* and *habeo*, which were joined at a later stage by others formed with the verb *teneo*, given its semantic and functional ties with *habeo*. The periphrases with *habere* did not require any kind of complementiser between their

³ For the scope of the notion of *conflict site*, a key element of the comparative variationist framework that accounts for the way in which certain linguistic forms differ functionally or structurally in different languages or varieties, see Poplack and Meechan (1998: 132).

elements in classical Latin, but both in the later period of this language and later still in the proto-Romance era several prepositions began to appear. This development took place in order to highlight their modal meaning and differentiate it from the mere future temporality that was arising from the grammaticalisation of the ancient periphrasis (Gili Gaya, 1970: 112).

According to Yllera (1980: 100-101), the periphrases with *haber* had a broad modal value in the early days of Castilian Spanish and expressed both necessity and obligation in a general sense. They also indicated a number of more precise shades of meaning (moral obligation, attenuated necessity, obligation based on the law or custom, etc.), in some cases even alternating with *deber*. Lapesa (2000: 882) also discussed these obligative shades of meaning and added others with a basically prospective value, together with some more sporadic ones of a pleonastic nature, where neither obligation nor futurity seems to play any kind of role at all. Nevertheless, by the end of the mediaeval period these latter had disappeared, and the periphrasis was left for centuries as the dominant expression for stating obligative and, to a lesser extent, futural contents.

On the other hand, the appearance of *tener* as an auxiliary verb in distributions that are analogous to those formed with *haber* has been associated with the process of grammaticalisation and semantic neutralisation of the two verbs that was to take place as of the Middle Ages. In this respect, Hernández Díaz (2006) noted that the language change that occurred between the 12th and 16th centuries was determined by a number of (semantic, but also syntactic and pragmatic) factors. This lexical erosion of *haber* was undoubtedly furthered by its widespread usage as an auxiliary in the formation of compound tenses and its extension as an existential and impersonal verb. In any event, as of the Late Middle Ages the obligative periphrases with the auxiliary *tener*, as well as some other modal and futural uses, ended up becoming firmly consolidated in Castilian Spanish (mainly with the variant *tener de* in the case of the futural uses, Author and Author Y, in press). The same phenomenon also occurred in other Romance languages, as in some southern Italian dialects, Portuguese, Gallego, Astur-Leonés and even, albeit more sporadically, in Catalan (see below).

It has been observed, however, that the case of *tener que* is unusual among the personal infinitive verb periphrases due to the fact that it is the only construction in which the two parts are not joined by a preposition (Pountain, 2001). Moreover, Olbertz (1998: 250) stated that, in relation to this combination, there are two types of syntactic structures, exemplified by those in (5) and (6). Yet only in the first one is there what we could call a true periphrasis in the strictest sense, although the second -in which a direct object depending of *tener* is interspersed between the verb and the complementiser- may well have played a significant role in its origins (Gutiérrez, 1980; Gómez Torrego, 1988; Olbertz, 1998; Pountain, 2001; Sinner, 2003):

(5) Pues *tengo que haceros esta advertencia*: ya que tenéis tanto tiempo libre, ¿por qué no lo empleáis leyendo, escribiendo y haciendo cuentas? (*Francia no nos llamó*)

[Thus I have to give you this word of warning: since you have so much free time, why don't you put it to good use reading, writing and doing sums?]

(6) Pues *tengo esta advertencia que haceros...*

[Thus I have to give you this word of warning...]

The history of the Spanish syntax is at the same time the history of an old-age change in progress by which *tener que* have been replacing *haber de* in modal obligative contexts. At first glance, this change can be noticed after comparing the frequencies of use of both periphrases in some diachronic macro-corpus. Thus, in table

1 we show the figures found by López Izquierdo (2008: 793) in the 100 million words *Corpus del español* (Davis 2002) between the 15th and 20th centuries. Leaving aside the possibility that a number of no truly periphrastic combinations can have been accounted for in these counts, the figures speak for themselves and confirm the existence of an important evolution in this grammatical paradigm. At the same time, the table reveals that the variation between the two periphrases is very stable until the 18th century, and only from the 19th century seems to move forward. Nevertheless, in this period *haber de* continues to be prevalent, in contrast to what will happen in the following decades of the 20th century, when things will change radically.⁴

	15th		16th		17th		18th		19th		20th	
	N	%	N	%	N	%	N	%	N	%	N	%
<i>haber de</i>	2339	86.5	17643	90	17329	90.9	4963	85.6	8458	76	1492	13
<i>tener de/que</i>	366	13.5	1915	10	1736	9.1	820	14.1	2680	24	9894	87

Table 1: Frequencies of use of *Haber de* / *tener que* + infinitive by centuries in the *Corpus del español* (Source: López Izquierdo, 2008: 793)

Our own figures, obtained in a time span of five centuries (16th to 20th) from a 3 and half million word corpus of immediacy texts, initially show some differences in these magnitudes. Thus, after comparing Tables 1 and 2 it seems that, in the discursive traditions closer to orality, the ratio of *haber de* in the nineteenth century is significantly higher than in other more formal traditions mostly represented in canonical macro-corpus at use (24% vs. 42.6%). Nevertheless, with all these differences of degree in mind, both tables agree on the stabilization of the change in the early centuries, as well as in the abrupt sorpasso of *haber de* in the 20th century. At the same time, a variationist comparatist analysis of the last two centuries allowed us to confirm how severe this change has been, both in general terms as in virtually all the linguistic contexts analyzed (Author and Author X, b)

	16th		18th		19th		20th	
	N	%	N	%	N	%	N	%
<i>haber de</i>	1584	91.5	693	89	162	57.4	272	26.1
<i>tener que/de</i>	168	9.5	86	11	120	42.6	771	73.9

Table 2: Frequencies of use of *Haber de* / *tener de/que* + infinitive by centuries in a corpus of 3 and million corpus of immediacy texts

However, a careful analysis of the linguistic envelop of the variation allows us to confirm the existence of certain loci in the grammar system that still exert some favourable action in the selection of *haber de* in the 20th century. An even more significant finding is that both the factor groups and the hierarchy of constraints that operated in this century were identical to those that have functioned in the past, Nevertheless, with odd exceptions, some generalized decrease in the explanatory power of these factor groups is observed, as well as some signs of lexicalization with a chunk of few lexical contexts, such as *ser* and several verba dicendi, mainly when they are used with a phatic meaning and in some specific contexts (see section x below).

⁴ The figures reported by Martínez Díaz (2003), obtained from several written corpora, point in the same direction. Hence, in the 18th and 19th centuries *haber de* still dominated over *tener que* in all kinds of texts. In the 20th century, however, the proportions are clearly inverted

As a result of a long process of grammaticalization, sharply accelerated in the first half of the 20th century, the periphrasis *haber de* + infinitive has been relegated to some restricted areas of the grammar and the lexicon in front of *tener que*, that seems to have almost definitely won the battle in this age-long case of variation and change. What we are considering now is whether this framework is distributed regularly from a dialectal point of view or, on the contrary, some differences can be observed among the Spanish varieties of several peninsular regions. And, if that latter scenery comes true, whether these differences can be related to the existence of structural points of conflict between the Spanish and other Romance languages in which similar forms -but not always similar functions- are present. In the next paragraph we will see in detail some of these (dis)similarities, leaving for a later section (section 5) its theoretical implications in the light of the results obtained in the variationist study.

2.2 *Tener que* and *haber de* in other Peninsular languages

Two Romance languages that share modal infinitive verb periphrases with Castilian are Gallego and Astur-Leonés. In reference to the first, Rojo (1974) pointed out that the fundamental value of the constructions with *haber* (which are possible with the prepositions *de* and *a*, although they also occur alone) is the expression of futurity. This is especially the case in the present and imperfect indicative, where this periphrastic expression even surpasses other prospective variants such as the morphological future (in this same sense, see Kabatek, 1996; Álvarez, 1983; Álvarez et al., 1998). Nevertheless, and as pointed out by Sinner (2003: 2001), sometimes it is difficult to distinguish between a futural shade of meaning and mere obligation, since they are notions that are very often linked together. In fact, for Álvarez et al. (1998: 406) “it can be said that in some cases temporality is the most important thing and in others obligation predominates” (our translation). Moreover, these authors also highlight the existence of epistemic senses with the periphrasis that can also be found in Castilian Spanish: “Ese rapaz *ha de andar* polos quince anos” [That guy has to be fifteen years old]. Yet, in a study on linguistic attitudes, Kabatek (1996: 136) observed how the dominant Gallego-speaking population tended to reject the modal, not futural, values more often than speakers who were more influenced by Castilian Spanish. Furthermore, according to some linguists this temporal mark left by the Gallego *haber de* makes itself felt in the Spanish spoken in that community, although apart from some more or less well-founded intuitions, the data available are fragmented and not very representative (Iglesias, 1969; Rojo, 1974: 83; García, 1976; Seco, 1986: 214; Rojo, 2004: 1095). Following an analysis of several interviews carried out in Santiago de Compostela, Sinner (2003) noted that the futural value of that Galician sample is not very representative, although it still appears occasionally, unlike the case of other control groups (inhabitants of Madrid and Catalonia) in which it never occurs.

In contrast, the periphrases with *tener* in Gallego fall squarely within the realm of the deontic modality (Rojo, 1974: 70), thus showing themselves to be much closer to the corresponding Castilian ones.

The semantic distribution of these periphrases is similar in another north-western Romance language, namely Astur-Leonés. Thus the *Academia de la Llingua Asturiana* (2001) draws our attention to the optional alternation of the periphrases with *haber* with the prepositions *a* and *de* to express either futurity or obligation, even with specific uses that are close to those found in Gallego (Kabatek, 1996: 136), such as the future of imminence (“en tal peligru me vi qu’hubi morrer”) [I saw myself in such danger that I wanted to die]. And the combinations with *tener* (also with the alternating

complementiser *de* and *que*, as in Gallego) are stated as being used in an obligative sense in which there is also “a clear futural shade of meaning” (our translation).

Catalan, on the other hand, offers a different scenario. Although in the mediaeval period this language contained several examples of *tenir de* + infinitive, which alternated within the sphere of modality with the more usual forms *haver de* and *haver a*, they soon disappeared. Hence, the far more recent uses of *tenir que* + infinitive have been condemned time and again as an unacceptable syntactic Castilianism that does not form part of the normative language. Thus, Badia i Margarit (1985: 391) noted how, in addition to other interferences in this area of modality: “... other *Castilianisms* are common in obligative constructions: the Castilian ‘tengo que decir’ is translated by *tinc que dir* [unacceptable], rather than *haug* (or *he*) *de dir* or *cal que jo digui*” (in the same sense, see Marvà, 1983; Payrató, 1985).

This absence of *tener que* in the periphrastic paradigm of Catalan has been proposed as being the reason underlying the hyper-representation of *haber de* in the Castilian Spanish of Catalonia and other Catalan-speaking regions, with respect to other areas of the peninsula (Wesch, 1997; Hernández García, 1998; Sinner, 2003, 2004; Author, 2004; Martínez Díaz, 2002, 2003, 2008; Sinner and Wesch, 2008). The empirical results of this presence, however, are far from being unanimous and vary from one study to another. Thus, Sinner (2003) noted that on going from the written language to a corpus of oral interviews the average uses of *haber de* drop from a high figure of 43.3% to just 4%. Martínez Díaz (2008), on the other hand, raises these frequencies to 17% (N = 9) in the *Corpus del español conversacional de Barcelona y su área metropolitana*, although these also contrast with the much higher figures for *tener que* (83%; N = 43). Yet, other authors such as Wesch (1994: 173-174) and Hernández (1998: 578) detected higher figures in other oral corpora. Furthermore, the use of *haber de* in Catalonia today could have a far from negligible identity-based component or, at least, be a faithful reflection of some of the notable outcomes of the linguistic policies of recent years. Thus, as stated by Martínez Díaz (2003: 690): “the frequency of use of the periphrasis ‘*haber de* + infinitive’ in the Spanish of Catalonia is not the same in all informants. Knowledge of the normative variety of the Catalan language interferes with the subvariety of Spanish but essentially in those speakers who have been educated in Catalan or who have learnt Catalan at school” (our translation) (for other phenomena in the same line in these regions of the peninsula, see Vann, 2002, and Author, 2008 a).

In sum, from what we have seen above it can be deduced that different Romance languages on the peninsula present different points of contact, but also points of structural conflict within the paradigms of personal infinitive verb periphrases, which could justify differences in the use made of them in the Spanish of different dialectal areas. Thus, while all the languages present infinitive verb periphrases with the verb *haber* in their paradigm (sometimes with different alternating complementisers), the same cannot be said for *tener*, which is absent from Catalan grammar except for some recent uses that are clearly syntactic calques from Castilian Spanish. Furthermore, the uses of *haber* in the north-western languages, such as Gallego and Astur-Leonés, do represent obligative values but these are less idiosyncratic than the futural ones, which are also shared by Spanish and Catalan, although to a lesser degree.

To what extent can these structural differences condition the uses of the periphrases *haber de* and *tener que* + infinitive in the different dialectal areas of the Spanish language? How have those constraint factors evolved over the last two centuries? To answer these questions we will conduct a study that can be included within the principles and methods of historical sociolinguistics (Conde Silvestre, 2007;

Nevalainen and Raumolin-Brunberg, 2003, 2012) and more specifically in that which addresses the study of variation phenomenon from a variationist and comparatist perspective (Tagliamonte, 2012; Poplack et al., 2012).

3. CORPUS AND METHODOLOGY

As part of a research project focused on the diachronic study of modal infinitive verb periphrases from classical Spanish until the present (see footnote 1), for this study we compiled a corpus made up of texts that are close to the pole of communicative immediacy (Oesterreicher, 2004), written by individuals from different social and diatopic origins. In order to ensure a higher degree of dialectal congruence in the data, we limited the analysis to texts written by Spaniards or by individuals born outside Spain, but who had spent most of their life in this country. In addition, the texts offer a number of different registers, ranging from the most personal or informal matters at one end of the spectrum, through different intermediate degrees of intimacy to others of a clearly less private nature at the other (for further details, see subsection 4.9).

Oesterreicher (2004) considers that every discourse are placed at a particular point in a conceptional axis between two extremes represented by the pole of 'spoken' or 'communicative immediacy', on the one hand, and the pole of communicative immediacy, on the other. In this sense, Oesterreicher (2004:734) says: “a private letter written by a half-educated man or the statements made by a humble person in the proceedings of a tribunal, may introduce forms that are relatively close to the pole of immediacy”. At the end, the private nature of many of this kind of texts has proved to be an especially attractive source for the study of the common language in earlier periods of the history of the language for which no oral testimonials have survived (Elpass, 2012).

For the 20th century, this corpus contains 24 works, mainly collections of private letters although there are also several autobiographical texts (account book, memoirs, diaries, etc.). Altogether it contains the writings of over 350 different speakers and 695,090 words. Data for the 19th century, on the other hand, are based on 28 texts of the same type written by 250 different authors, with a total of 490,014 words. This figure is clearly lower than that obtained in the next century, which explains, albeit only partially, a significantly smaller degree of representativeness of the sample.

All the occurrences of the two variants in the corpus were selected using a concordance program (*Wordsmith* v.4) and they were then coded according to over 20 factors of a linguistic, stylistic and social nature, the details of which have already appeared elsewhere (Author and Author X, *a* and *b*). Due to space restraints, in the next section we will only deal with the factors selected as significant by the logistic regression program Goldvarb 3.0. As is well known, with this statistical program not only is it possible to calculate the differences in frequency among the different variants and their contexts but also, and more importantly, the degree of significance and the explanatory hierarchy of the factors being analysed, when all of them are considered at the same time (Tagliamonte, 2012). This also allows relations of (inter)dependence and interaction among different factors to be discovered, which is far more difficult with a merely descriptive statistic.

4. RESULTS AND ANALYSIS

4.1 *General data*

The total number of periphrases was 1326, although they were distributed in a very irregular manner, since 78% (N = 1044) were from the first six decades of the 20th century, while only the remaining 22% came from the 19th century (N = 282). The fact that the corpus from this century was considerably smaller (about 200,000 words, as we have seen) does not in itself justify a difference which we have not been able to find a plausible explanation for.

Of these periphrases, 831 (70.8%) contain *tener que*,⁵ as opposed to 342 (29.2%) with *haber de*. Yet, this distribution also varies notably on the diachronic axis, as can be seen in the graph below, which shows the uses of *haber de* in four different periods on that time continuum: 19th century: 1st and 2nd half, and 20th century: 1st third and 2nd third.

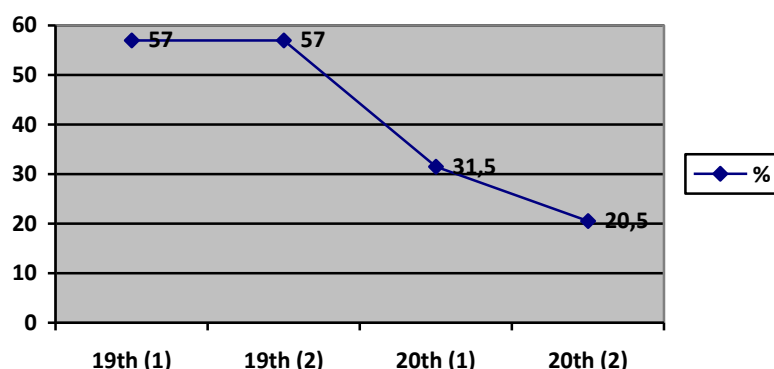


Figure 1: Frequency of usage of *haber de* + infinitive in different periods of the 19th and 20th centuries in a corpus of communicative immediacy (%)

Although to a lesser extent, the differences in frequency can also be observed on the diatopic plane, a fact that can be evaluated using the different dialectal sources of the texts in the corpus. For the purposes of this study, the corpus was initially divided into:

a) texts written by individuals from Spanish monolingual communities, in which Castilian is the only language used in day-to-day life, and

b) texts from bilingual regions, where Spanish has coexisted for centuries alongside another autochthonous language, which conditions a number of linguistic usages (Author, 2004, 2008 a; Rojo, 2004; Sinner, 2004). Nevertheless, of these latter areas, two have been isolated independently with the aim of performing a more accurate analysis of the potential influence that language contact has on the variation phenomenon we are dealing with here. Thus, we distinguish between:

b.1) varieties from the Catalan-speaking territories (today's autonomous communities of Catalonia, the Valencian Region and the Balearic Isles, although there are no testimonies from this last area in the corpus), whose autochthonous language shares multiple uses of *haber de* with Castilian Spanish but lacks a second alternative periphrasis (*tener que*), unless we can count a recent syntactic Castilianism that the normative grammar strongly rejects;

⁵ Within this group we also included 15 occurrences of periphrases with *tener* that use the preposition *de* as a complementiser (*tener de* + infinitive). Despite its notable vitality in times gone by (Yllera, 1980; Author and Author Y), use of this periphrasis in modern Spanish is now limited to some very restricted dialectal uses.

b.2) north-western bilingual regions, more specifically Galicia and Asturias, whose Romance languages include the two periphrastic forms in their verbal repertoire, although with some idiosyncratic semantic values that do not always coincide with those of Castilian Spanish, as we saw earlier (see section 2).

The initial results of the analysis concerning this dialectal distribution show that (see Figure 2):

a) as expected, the Catalan-speaking regions make greater use of the periphrasis that is common to both languages, *haber de* (N = 158; 41%) than the monolingual territories (N = 101; 30%). Since the expressive habits of Catalan-speakers contain only one periphrasis in this grammatical domain, it seems logical to expect that when they express themselves in the other language used in the community they resort to that verbal construction more often than speakers from monolingual regions.

b) this is not the case, however, in the other bilingual areas. In fact exactly the opposite happens, the districts of Galicia and Asturias being the regions where *haber de* is used with the least frequency in absolute terms (N = 83; 18%).

c) as in the general data outlined above, important differences are also seen here on the time axis. Hence, whereas the general behaviour of the three dialectal groups is practically identical in the 19th century, the differences get wider as we go further into the 20th century, when *haber de* becomes increasingly less prominent in favour of *tener que*.

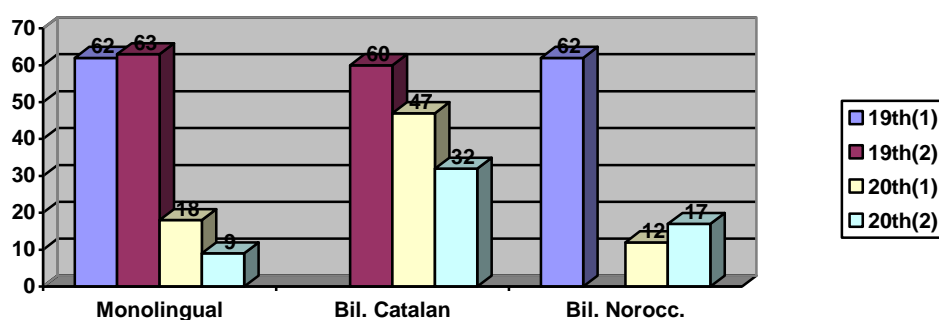


Figure 2: Frequency of usage of *haber de* + infinitive by periods and dialectal origin of writers (%) (NB. They are not data available for the Catalan groups (1st half 19th century) and Galician-Asturian (2nd half 20th Century)

As can be seen in Graph 2, that loss of prominence is especially intense in the monolingual group, where the selection of *haber de* falls sharply from frequencies above 60% in the 19th century to a little over 8% in the second third of the next century. Nevertheless, this decline is also just as significant in the north-western bilingual areas, whose distributional profile veers sharply away from the far gentler change undergone by the areas within the Catalan-speaking territories.

4.2. Temporal uses of the periphrasis

One preliminary hypothesis to account for the initially unexpected behaviour of the Gallego-Asturiano bilingual group leads us to consider the potential influence of some of the futural uses of the periphrases. Together with the modal contents, in the corpus there are also others in which the periphrases have no modal value whatsoever, or they are now very diluted in favour of the expression of merely temporal shades of meaning.

It is true that the actual modal meaning often implicitly contains a temporal value, since most of these periphrases possess an ingressive or inchoative aspectual value that, by nature, points towards futurity. Nevertheless, the speaker sometimes leaves the modal aspect to one side and his or her utterance appears to be intended as just a means to formulate facts situated in some future time, as can be seen in the following examples:

(7) ... y esperando ocasión oportuna para pagarle esta deuda de gratitud. Vamos, que se *ha de alegrar* con la nueva que va a oír (*Cartas de San Enrique de Ossó*)
[... and awaiting the occasion to pay him this debt of gratitude. I mean, he has to be pleased with the new one he is going to hear]

(8) Tú, María, no te muevas de ahí mientras no vaya yo; si no, no *he de ir a verte* aunque estés en Pamplona (*Once cartas de mi padre*)
[You, Maria, don't you move from there while I am not there; otherwise, I shall not go and see you even though you are in Pamplona]

At this point it should be remembered that in old and classical Castilian Spanish these futural values were relatively important. Thus, according to some studies based on literary texts from different periods, the periphrasis *haber de* + infinitive went from being used with its future value in 17% of cases in Cervantes' *Entremeses* and 18% in Lope de Vega's comedies to just 1% in the drama of the second half of the 20th century (Sáez Godoy, 1968). Furthermore, some authors have drawn attention to the fact that such values have survived to a greater extent in American speech forms than in European ones (Kany, 1969; Steel, 1982; De Bruyne, 1993; Westmoreland, 1997). The temporal non-modal meanings also appear, although with less frequency, in periphrases with *tener*, and more specifically in those that employ the preposition (*tener de*) as the complementiser between the auxiliary and the main verb (Yllera, 1980; Author 1 and Author Y). Yet, our data show that, at least in the last two centuries, the future is almost categorically associated with *haber de*, and the periphrases with *tener* are excluded from this semantic sphere.

Now, as we saw earlier (see section 2), the futural uses of the periphrases that utilise the verb *haber* are not restricted to Spanish. In fact, they are especially characteristic in other Romance languages, such as European Portuguese (although not so in the Brazilian variety, where it is considered archaic), as well as several Spanish languages, like Gallego and Astur-Leonés. Accordingly, if the influence of language contact is relevant on this point, we would expect there to be a difference favouring the north-western speech forms over the other two groups, i.e. both the monolinguals and – even more – the bilinguals. However, the results do not endorse this dialectal differentiation, since the temporal uses are distributed almost categorically over the three groups: Monolinguals (100%; N = 33), Gallego-Asturiano (97%; N = 28) and Catalan (92%; N = 34). Yet, where these differences are seen to be significant is, again, on the diachronic axis. Hence, these merely prospective values drop from a far from negligible 21% in the 19th century to just 5% in the 20th century, but again with very small differences between the three groups: Monolinguals (19th: 20.5%; 20th: 4.5%), Gallego-Asturiano (19th: 23%; 20th: 5.3%) and Catalan (19th: 23.8%; 20th: 5.5%).

4.3 Variation in the sphere of modality

4.3.1 Epistemic and expressive modal senses

So, discarding the unequal use of the periphrasis with a futural value across the different regions of Spain, our interest now focuses on analysing its modal uses, which represent 91% of all the occurrences of the variable. Some of these usages, such as (9) and (10), express epistemic contents, related with the notions of probability or conjecture, and which can be neutralised with other modal infinitive verb periphrases, this time with the verb *deber* (Gómez Torrego, 1999: 3353):

(9) ... conque sale por consecuencia que mi ermana *havia de tener* oculto lo que menos ese caudal (*Muestra documental del castellano norteño*)
[...which in consequence means that my sister should have at least that amount had hidden away]

(10) ... el cartero leía las cartas y no faltaban más que tres sin haber aparecido la tuya aún, aunque no perdía la confianza de que *tenía que estar* allí por ser ya hoy jueves (*Once cartas*)
[... the postman read the letters and when there were only three left yours had still not appeared, although I continued to trust that it had to be there because today is Thursday]

Thus, in the early nineteenth century, the Basque writer of a letter addressed to a friend expressed in (9) an assumption about the amount of money that his sister must have hidden. Meanwhile, in (10), the Navarre soldier who during the Spanish civil war writes to his wife, tells her how, despite having no certainty, he had the presentiment that her long-awaited letter had arrived that day, because "today it's Thursday" (then, the usual day to receive mail in the front).

Nevertheless, the level of representativeness of these epistemic modal shades of meaning is low (5%), both for *haber de* (N = 42) and, even more so, for *tener que* (N = 22), with figures that are a long way from those obtained by *deber (de) + infinitive*. In this latter case, for the same dates as those addressed in this study the figures are three times (N = 212) those obtained here (Author and Author Y, in press).

Lower still is the degree of representativeness of these verbal constructions to express different shades of meaning, such as surprise, indignation, obviousness, etc., which endow them with a particular expressive emphasis (Gómez Torrego, 1999: 3356). This is the case, for instance, of (11), where the author of a letter writes to someone about the contradiction that would result from acting now in a different way to how he or she did in the past. Or in (12), where the sender of another letter writes to tell a nephew the reasons why he or she is so fond of him. Yet, there are 13 examples of these shades of meaning for *haber de* versus just 4 for *tener que*:

(11) Si yo me acusé de dicha falta ¿cómo *he de presentar* descargos? (*Carta familiar de D. José Butrón*)
[If I accused myself of such a fault, how am I to present a case for the defence?]

(12) ... para mí eres mucho más que un sobrino pero porque te ayudé a criar, te tuve muchas veces en mis brazos y entonces cómo no *tengo que tenerte* cariño (*As cartas do destino*)
[... you are far more than a nephew to me but because I helped bring you up, I held you many times in my arms and so how can I not be fond of you]

The scarcity of these epistemic and expressive uses of the periphrases contrasts with the deontic values, which are related with notions of obligation or necessity, and where most of the periphrastic uses that interest us here are concentrated (85%; N = 1115). And this is the section where the differences among the areas used in the corpus can be seen most clearly, with the Catalan-speaking regions (34%) doubling the percentages of *haber de* of the monolingual areas (18%) and with values three times the indices obtained in the north-western bilingual territories (10%). But what is the tenor of these

differences? do they affect the envelop of variation, that is, the linguistic and extralinguistic conditioning of periphrases? Or to put it another way: is the grammar underlying this variation different in one or any of these dialectal areas?

To answer these questions, in the following we present the findings from three independent variable rules analyses performed for each of these regions. Table 3 shows the results of these analyses, which take into account the incidence on the variation of a number of different linguistic, stylistic and social factors when they are all considered at the same time. As we shall see, despite the above-mentioned differences in frequency, there is a notable degree of consistency among the results from the three groups. This would confirm the fact that, except for the odd exception (which as we shall see are also revealing), the processes of variation and language change in all of them are conditioned by similar factors.

Factors	Monolingual regions			Bilingual Catalan-speaking areas			Bilingual Northwestern-speaking areas		
	P.	%	N	P.	%	N	P.	%	N
Century									
19 th	.81	62.3	71	.78	60.2	53	.86	57.7	15
20 th	.31	13.6	30	.41	35.8	105	.37	15.8	68
<i>Range</i>	<i>50</i>			<i>37</i>			<i>49</i>		
Manner of verbal action									
Estatic verbs	.29	25.7	36	.60	47.5	67	.48	21.6	33
Verba dicendi	.74	37.2	16	.78	46.6	27	.76	21.4	12
Verbs of movement	.45	23.3	14	.17	24.6	14	.24	12.5	7
Other verbs	.74	37.4	34	.42	40	50	.51	16.1	31
<i>Range</i>	<i>45</i>			<i>61</i>			<i>52</i>		
Modal senses									
Internal oblig.	.26	13.3	6	.34	42.4	25	.72	20	14
External oblig.	.47	18.1	30	.58	33.7	64	.39	8.2	20
Necessity/Convenience	.44	25.4	16	.32	30.8	24	.45	9.2	8
Other	.94	59.3	16	.86	72.2	13	.92	61.9	13
<i>Range</i>	<i>68</i>			<i>52</i>			<i>53</i>		
Verb tense and mood									
Present indicative	.59	34.6	71	.71	53.5	114	.54	21.4	62
Imperf. Indicative	.65	52.5	21	.39	37	10	.72	21.2	11
Future indicative	.09	5.9	1	.61	54.8	23	.41	12	3
Other	.29	11	8	.11	11.1	10	.21	7.8	7
<i>Range</i>	<i>56</i>			<i>60</i>			<i>51</i>		
Level of (im)personalization									
Active sentences	.45	27.6	88	.45	38.5	135		17.7	80
Pasives-impersonal sentences	.97	81.2	13	.86	76.7	23		75	3
<i>Range</i>	<i>53</i>			<i>41</i>			--		
Type of clause									
Subordinated		34.1		.61	45.5	102	.62	25.6	69
Other		24		.35	35.7	56	.32	7.5	14
<i>Range</i>	--			<i>26</i>			<i>30</i>		
Style									
Intimate letters	.39	20.7	43	.09	9.8	12		16.3	65
Distant letters	.75	61	25	.61	53.4	62		16.7	1
Non-epistolary	.62	40.2	33	.83	59.2	84		32.7	17

(autobiographical) texts									
<i>Range</i>	36			74			--		
Socio-cultural level									
Low		11.2	14	.23	16.7	6	.38	10	30
Middle		27.4	31	.35	33.3	11	.72	34	38
High		54.9	50	.56	45.3	141	.74	32	10
<i>Range</i>	--			33			36		
Age									
Young cohorts		24.2		.17	26.5	22		18.9	63
Adult cohorts		34.4		.61	48.7	135		12.5	12
<i>Range</i>	--			44			--		

Table 3: Three independent variable rule analysis for the selection of *haber de* + infinitive in three dialect areas of Peninsular Spanish (Goldvarb 3.0)

4.3.2 The expression of deontic modality

As pointed out earlier, the most frequent uses of *haber de* and *tener que* occur within the sphere of the deontic modality. Different authors have attempted to catalogue its shades of meaning, but this is no simple task and sometimes leads to rather circular arguments (Keniston, 1937; Yllera, 1980; Olbertz, 1998, Gómez Torrego, 1988, 1999; Fernández, 1999; García Fernández, 2006; López Izquierdo, 2008, Martínez Díaz, 2008). In an attempt to get as far away from excessive subjectivity as possible, we have placed the deontic contents at different points on an imaginary axis, on which we combine two parameters that often appear in the literature, namely:

- a) the degree of obligation/necessity imposed; and
- b) the agent that imposes that obligation or necessity.

On crossing these two parameters, the following main values are obtained:

1. *Subjective (internal) necessity or obligation.* In this case we are dealing with a sense of duty based on inner conviction or on the subject's determination or intention owing to motives that may be of a religious, ethical or philosophical nature or that perhaps arise from gratitude, respect or any other kind of motivation. It is thus based on the subjective convictions or the desire of an agent, and it is therefore this latter more than anyone else who feels the need to fulfil it, which brings this periphrasis closer to those of a volitional nature (Roca Pons 1980: 73; Yllera 1980: 114). Those in (13) and (14) are representative examples:

(13) ...pero es necesario consolarse porque es una carrera que todos *tenemos que pasar*...
(*Historias de América*)

[... but we must comfort ourselves because it is a race we all have to finish]

(14) Creo que la política de ahora no *ha de ser* de engaños ni es cuestión de forjarnos vanas ilusiones que después la realidad de los hechos ha de desvanecer. (*Un catalanófilo de Madrid*)

[I believe the politics of today must not be about deceiving nor is it a question of building up illusions in vain for them to be later dissipated by the reality of the facts]

2. *Agent-oriented or external obligation.* Contents involving obligation, unavoidable necessity or imperative and coercive advisability are of an *external* nature in relation to the agent of the action described by the verb. Hence, we are dealing with

directive statements, among which different possible shades of meaning can be distinguished, such as: a) obligation imposed by norm, agreement, social convention, legal code, etc.; b) mandate or external command to compel someone to perform an action; c) obligation imposed by external circumstances, that is, the idea of an agent beyond the subject's will; and d) inevitability, where the speaker is so certain about the idea projected into the future that its realisation is considered necessary or unavoidable. The following are some illustrative examples of each of these values:

(15) ... para el efecto *tendrás que entenderte* con el comandante general de marina (*Historias de América*)

[... to this effect you will have to reach an agreement with the Commander-in-Chief of the navy]

(16) Si llegamos a tener dichos escritos, al instante se *han de imprimir* (*Epistolario José de Azara*).

[If we eventually get hold of those writings, they must be printed immediately]

(17) ...pues si te hallases aquí no te faltaría nada a la vera de tu hermano que *tiene que servirse* de gente extraña, pudiendo estar juntos (*Cartas de emigrantes escritas desde Cuba*)

[...if you were here, you would want have everything you need here next to your brother who has to take on strangers, when you could be together]

(18) Sofia siempre la misma con un Lujo enorme y Paboneándose por aquí y por allá. S, Candido y la mujer por aquí rrecojiendo la cosecha aprovechado hasta los Botones del suelo sin pensar que *han de morirse* (*Una familia y un océano*)

[Sofia the same as ever exhibiting vast amounts of luxury and swaggering around showing herself off. S., Candido and his wife were here harvesting the crops and even collecting buttons off the ground without thinking that they have to die]

3. *Necessity or advisability*, considered by the speaker, and therefore with far less coercive power than that expressed in the examples above. Thus, in contrast to (17), where the sender of the letter regrets having to engage strangers in the distant land he has emigrated to, in (19) and (20) the authors express their conviction regarding the advisability or necessity of certain states of affairs occurring:

(19) Allí mismo escribí otro volumen que titulo "Cinco hombres", impresiones sobre Pablo Iglesias, Jaime Vera, Tomás Meabe, Largo Caballero y Julián Besteiro. Son a la vez crítica de un libro de cada uno de ellos. *He de completarlo* con algo más de lectura. (*Dramas de refugiados*)

[Right there I wrote another volume entitled "Five men", impressions of Pablo Iglesias, Jaime Vera, Tomás Meabe, Largo Caballero and Julián Besteiro. At the same time they are a review of a book by each one of them. I must finish it with a little more reading]

(20) ... y si no se puede trabajar en el campo que es donde *tiene que salir* la riqueza del país, estamos de más, apaga y vámonos (*Cartas desde América*)

[... and if we can't work in the fields, which is where the country's wealth has to come from, we aren't needed, then let's get out of here]

Lastly, the analysis of this modal factor is completed with an additional group containing the other non-deontic modal values (epistemic and expressive), which we referred to earlier (see subsection 4.3).

The modal factor is selected as significant by the three dialectal groups. The contexts that most favour *haber de* are the non-deontic modal senses, with very high P values in the three cases, which show that all the peninsular speech forms display a substantial preference for this periphrasis. In the same way, in the three dialectal areas there is a common resistance to the selection of this periphrasis in the contexts of necessity or advisability. In contrast, certain differences can be observed as regards

obligation, with results that partially disagree with each other. Hence, internal obligation clearly disfavors *haber de* in the monolingual (.18) and Catalan (.32) groups, but favors it in the Gallego-Asturiano group (.68). On the other hand, the Catalan group diverges on agent-oriented or external obligations, with a slightly positive influence (.53) that contrasts with the negative effect in the other two groups.

4.4 *Manner of verbal action*

The significance of this factor in infinitive verb periphrases at different moments in the history of the Spanish language has been noted in several previous works (Balasch 2008, 2012; Author and Author X *a* and *b*; Author and Author Z). This same significance can now be seen in the alternation between *haber de* and *tener que* during the 19th and 20th centuries, a period in which it is selected as significant in the three dialectal groups under study. Moreover, the three groups fully agree as regards which verbs (dis)favour each variant the most. Thus, the *verba dicendi* are the contexts that most favour the selection of these periphrases with *haber de*, with surprisingly similar P values in the three dialectal areas. As we revealed in a previous study (Author and Author X *a*, in press), some kind of lexicalization can be observed with these verbs, especially in the 20th century data, to the point that many of the more frequent main verbs accompanying these periphrases correspond to such *verba dicendi*. In descent order of frequency, this is the case of *decir* [to say], *saber* [to know], *confesar* [to confess], *reconocer* [to recognize], *expresar* [to express], *escribir* [to write], *agradecer* [thanks] and *juzgar* [judge]. Especially significant is the case of *decir* [to say], that also figured among the most frequent verbs in the past but at considerably lower rates, as can be accounted for by the fact that it has passed from 13th position in the 19th century to the 2nd one a century later. Moreover, a more detailed analysis of these verbs show how many of its periphrastic uses appear in what we described above as "phatic" context, as in (21), in which the speaker seems to rely on his or her necessity or desire to "enter in communication" (Gómez Manzano 1992: 160; Gómez Torrego, 1999: 3354), a particularly favourable context in the epistolary genre.

(21) Por lo que a nuestro querido y llorado José María se refiere, *he de decirles* que lo he tenido muy presente en la Santa Misa (*Cartas de un requeté*)

[As our beloved José María is concerned, I must tell you that he had a strong presence in the Mass officiated by me]

Now, we are in a position to confirm that this lexicalization effect, that has been related to the loss of productivity of a receding variant in advanced stages of language change (cf. Poplack & Dion 2009, Elsig 2009: 19) is very congruently distributed in our corpus, as it can be deduced from the high and similar factor weight obtained by this constraint in all dialectal areas.

This same congruency can be perceived, on the opposite side, among the verbs of movement, which –with the odd exception of *ir* [to go]– are by far the main verbs that least frequently combine with the auxiliary *haber* (and, in contrast, those that do so most often with *tener*). Now, although it is true that the figures are low in all cases, they are especially so in the Catalan group (P. 17), which explains why the explanatory range of this factor in this dialectal area is higher (.61) than in the others. The behaviour of the stative verbs also differs in the Catalan case, with a positive influence (.60) that does not appear in the other two groups.

4.5 *Verb tense and mood*

We find ourselves before a new significant factor, both in the sample as a whole and in each of the dialectal areas that were analysed.

The first thing that should be noted about this syntactic factor is the important unbalance offered by the different conjugation paradigms in the sample. Thus, the present indicative (i.e. *tenemos que/hemos de*) is a hyper-represented form in all the groups (with percentages above 60%), followed at a considerable distance by the imperfect subjunctive (i.e. *teníamos que/habíamos de*) (with values close to 10%), the future simple (i.e. *tendremos que/habremos de*) (7%) and, even further behind, the other eleven verb forms. Of these, only 5% (N = 62) are in the subjunctive mood (i.e. *tengamos que/hayamos de*), a percentage that is similar to that of periphrases in which the auxiliary verb has been replaced by a non-finite form (infinitive and gerund – there are no examples of participles) (i.e. *tener, teniendo que/haber, habiendo que*). The others are indicative forms. For the regression analysis, we reduced the eleven paradigms to four, those consisting of the present, imperfect and future indicative, on the one hand, and the rest of the forms, on the other. The latter were included in a single group not only because they were less represented in the corpus, but above all owing to the clear preference most of them displayed to associate with the *tener que* periphrasis revealed by the initial examination of frequencies.

The results of the multivariate analysis again offer notable amounts of congruence, which shows that this is not only one of the most significant factors for explaining the variation but also that its hierarchy is very similar in the three dialectal areas (range between 50-60). Moreover, some additional regularities were also discovered within the factor. Thus, in all the areas, the present indicative exerts a positive influence on the selection of *haber de*, although it is significantly more pronounced in the Catalan group (.71) than among the Monolingual (.59) and the Norwest Bilingual (.54). Similarly, there is a strong coincidence in the negative influence exerted upon the periphrasis by the minority paradigms included in the category ‘Others’, with very low values (below .30) in all cases.

Some divergences are also observed, however, in relation to the other two factors, where the Catalan group again displays a behaviour that is different to that of the other two. Hence, whereas the future forms in the Norwest Bilingual (.41), but about all, in the Monolingual (.09) group, disfavour the selection of *haber de*, the opposite happens in the Catalan speech forms (.61). And on the contrary, the imperfect indicative disfavours the periphrasis in these varieties of the Catalan-speaking territories (.38), in opposition to what occurs in the other peninsular regions (Monol.: .65; Norwest. Biling.: .72).

4.6 Other linguistic factors

This group also includes other factors whose explanatory significance has been corroborated by the regression analysis in two of the dialectal groups, but rejected in the third. The differences in frequency with regard to this last case, however, are in the same line as those observed in the other two, so we cannot therefore rule out a possible effect upon the result due to certain insufficiencies in the sample.

This latter is undoubtedly the case of a semantic factor such as the degree of semantic (im)personalisation of the sentence in which the periphrasis appears, and the one we use to test the hypothesis about the existence of a common deagentivising factor in the periphrases with *haber*. Thus, as stated by Stengaard (2003): “... by means of the periphrases with *aver*, the subject of the action expressed by the infinitive either loses

its possible role as subject-agent or reinforces its role as subject-receiver or patient involved in the verbal action in question” (our translation). This effect of meaning has to do with the semantics of the verb *haber*, which, in contrast to its counterpart *tener*, would represent a non-agentive or receptive possession where no control would be exerted by the subject over what is possessed (Seifert, 1930). This opposition was explained from a cognitive point of view by Garachana (1997) in terms of the prototypicality of possession, whereby *haber* underwent a semantic “emptying” (a figurative control over what is possessed) that *tener* did not experience. And in another study (Garachana and Rosenmeyer, 2011), the same author concluded that the agentivity of the subject is greater in constructions with *tener que* than in those formed with *haber de*.

To evaluate this idea we divided the sample into two groups, represented by a) active sentences, and b) passive (analytical and reflexive) and impersonal sentences, respectively. If, when considering the verb tense, we spoke of the hyper-representation of the present over the other forms, in this case the imbalance in the sample is even greater because active sentences account for about 90% in all the groups. Nevertheless, data from the analysis show how the factor is significant in at least two of the dialectal areas (monolingual and Catalan) and with the same distribution in both cases. Hence, we can see how passive and impersonal sentences, where the agent who performs the action is camouflaged, are the ones that most favour *haber de*, with P values close to unity. And although, on this occasion, the regression analysis has ruled out the statistical relevance of the factor in the north-western bilingual group, the same distribution profile can be seen within it, as shown by the frequency analysis (Active: 17.7%; Passive-impersonal: 75%).⁶

Something similar was observed in the analysis of a new syntactic factor, namely the distinction between the type of clause where the periphrasis appears, which led us to draw a distinction between subordinate and non-subordinate contexts. It has been said that subordination represents a structural locus that is not very favourable for processes of language change (Tarallo, 1989; Matsuda, 1993; Author, 2008 *b*), which leads us to wonder about the role it may be playing here. The results of the analysis largely support this hypothesis, as shown by a frequency distribution that follows the same pattern in all cases: subordinate clauses display a greater alliance with the traditional variant, *haber de*, than non-subordinate clauses. Further still, this distribution is selected as significant by the regression analysis in two of the dialectal areas (Catalan and north-western) and, moreover, with practically identical probabilistic values and ranges.

4.7 Extralinguistic factors

On the diachronic axis, the explanatory significance of the differences in the frequencies obtained in the 19th and 20th centuries, which we referred to earlier (see subsection 4.1), is confirmed. In the three dialectal areas, as one century gave way to the next there was a generalised slump both in the realisations of *haber de* and in the associated probabilistic values. Nevertheless, the hierarchy of this factor is greater in the monolingual (range 50) and north-western areas (49) than within the Catalan-speaking territories (range 37), which confirms the greater retention of the variant in this dialectal area.

⁶ Moreover, it could be possible that this time the factor has been eliminated as a result of the almost negligible representation of the passive and impersonal group, of which there are only four occurrences in the north-western group.

Together with the temporal axis, within this subsection of extralinguistic factors we also analysed the incidence of different stylistic and social parameters. The former were delimited by using the preliminary combination of two initial parameters, namely: 1) the main subject matter of the texts, a distinction being drawn in this respect between discourses of a more or less intimate nature; and 2) the distance in the relationship between the interlocutors. The result of this combination offers a stylistic continuum on which three points are located, with sufficient distance being left between them to allow a reliable comparison:

a) Letters containing private or intimate material between people who are closely related, whether this is due to kinship (most of the cases), friendship, love, and so forth.

b) Letters in which the predominant topic is not informal and in those which, moreover, there is a clear distance between the interlocutors on the axes of intimacy or solidarity.

c) Non-epistolary (autobiographical) texts (memoirs, diaries, court statements).⁷

For the phenomenon that we are dealing with here this stylistic differentiation is significant in the monolingual areas (range 36), although it is even more so in the Catalan-speaking regions, with an explanatory significance that is twice that of the previous group (range 74). In any event, both cases display an identical distributional profile: the less frequent selection of the *haber de* variant takes place within the more intimate and personal letters, rather than in the more formal and distant letters and autobiographical texts. The same differences between these latter cases (32.7%) and the intimate letters (16.3%) can be observed, moreover, in the north-western bilingual group, although in this case they are not statistically significant.⁸

This result is largely consistent with the fate undergone by the periphrasis with *haber* in contemporary Spanish, in which it has been confined in most cases to formal uses in the written language, in contrast to the more colloquial and everyday nature of *tener que*. In contrast to the letters with a higher level of spontaneity in their conception and execution, distant or formal letters are characterised by the opposite, and hence their greater association with *haber de*. And the same can be said of the autobiographical (non-epistolary) texts in which, although they involve subjects that can become quite intimate, the interactivity component that characterises letters is missing. This fact, as well as the possibility that the authors of these texts sight the idea that one day their memories may come to light and be read by other people, would necessarily have an effect on the degree of attention that lies at the base of stylistic differentiation (Labov, 1972).

Of the social factors considered at the outset, two display some degree of statistical significance: socio-cultural level and age. In accordance with the classifications most commonly used in sociolinguistics (Author, 2005; Conde Silvestre, 2007; Tagliamonte, 2012; Bergs, 2012), in this study the social spectrum has been divided up into three groups, which were characterised as low, medium and high. The significance of this sociolectal differentiation is confirmed by the regression analysis in the two bilingual dialectal groups and, moreover, with a very similar range. Taking into account the fact that the frequency profile is practically identical in the monolingual group, with the lowest strata leading the change towards the use of *tener que*, and the greater tendency of the higher socio-cultural levels to conserve the variant that had been dominant for

⁷ To make it easier to compare them, occurrences that did not fall prototypically within those three groups were not coded for this factor.

⁸ Very few data are available on the distant letters in this group, which probably must have had some effect on the lack of significance of the factor.

centuries (*haber de*), it seems that we are not only dealing with a factor group with a notable explanatory power but also have before us a possible change from below.

In addition to the promotion of these new variants by the socially less favoured sectors of society, one of the features characterising these changes from below is the fact they are often taken up and led by the younger generations (Author, 2005; Chambers, 1995; Labov, 2001; Tagliamonte, 2012), something that is also confirmed here. In order to configure this factor, in this study we carried out an emic rather than chronological interpretation (Chambers, 1995), given the difficulties involved in being able to determine the exact age of many of the speakers when they produced their writings. Accordingly, we have resorted to a binary classification, in which the speakers are classified in two groups, depending on what they were experiencing in their lives at the time of writing their texts. To do so, we used both what they said in those texts or the biographical information available from other sources (prologues, biographies, etc.) and – especially in the case of the letters – indirect data related with the actual communicative act itself.

With the sample thus composed from the genolectal point of view, a quantitative analysis revealed an identical pattern in two of the three dialectal areas (this time the exception was in the north-western area, where the differences between the groups are very small). Thus, both in the monolingual regions (34.4%) and in the Catalan-speaking territories (48.7%) older adults display a greater resistance to abandon the use of *haber de* than the younger speakers (24.2% and 26.5%, respectively). The latter, in contrast, lead the change favouring the alternative variant, *tener que*. However, the factor is only selected as significant in the Catalan area (range 44), a situation that is undoubtedly affected by the high proportion of uses of *haber de* among older adults, with figures reaching practically 50% (P. 61) in the period considered in this study. They are even higher in the 19th century (77%), although by no means negligible in the 20th (44%), which places this generational cohort at a notable distance from their counterparts in other Spanish regions.

Finally, the profile of change from below in the Catalan-speaking areas is endorsed by the interaction between the socio-cultural and generational factors. Although the sample differences between some subgroups now make it difficult to establish more accurate inferences, it is nevertheless revealing that, as can be seen in the graph below, in each and every one of the social sectors it is the young people who are leading the change favouring *tener que*. It is also interesting to note that this change is especially abrupt in the lower social classes, where no occurrences of the traditional variant *haber de* are found.

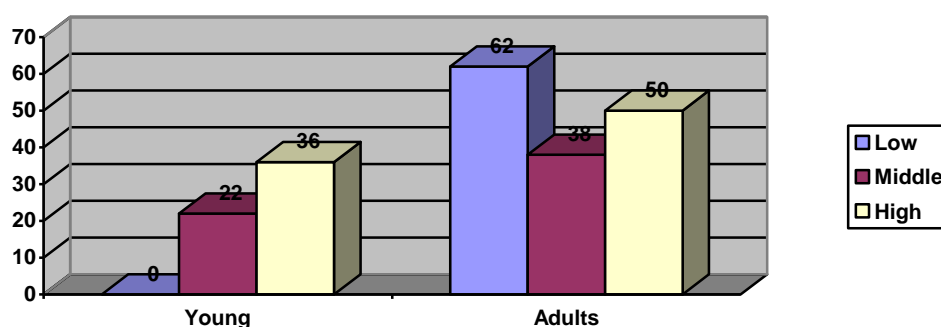


Figure 3: Frequency of usage of *haber de* + infinitive in the Catalan-speaking areas by generations and cultural levels (%)

5. DISCUSSION: THE SCOPE OF LINGUISTIC CONVERGENCE IN THE EVOLUTION OF MODAL PERIPHRASES IN PENINSULAR SPANISH

Over the last century, the personal infinitive verb periphrasis *haber de* + infinitive, which had clearly been the dominant form throughout the history of the Spanish language, lost most of its traditional usages in favour of *tener que*. The diachronic study conducted in this research project confirms this claim, and leaves no doubts about how this change gradually became faster as the 20th century progressed. Yet, in previous works we have seen how this process of variation and language change has been conditioned by several different structural, stylistic and social factors, whose explanatory significance, although weakened in most cases, has been conserved over time, at least up until the moment analysed here (the mid-20th century).

A review of the literature, however, alerts us to the possibility of such conditioning factors having some influence, and even a different fate, in some dialectal regions. This would essentially be due to the potential influence of the contact between Spanish and the other Romance languages alongside which the former has coexisted for centuries. It is not surprising, then, that a review of the respective grammars of these languages reveals the existence of structural points of conflict with Spanish that could potentially condition the uses of the periphrases in these bilingual communities. Thus, the verbal combinations with *tener que* are absent from Catalan grammar, unlike the cases of Castilian Spanish, Gallego or Astur-Leonés. Catalan and Spanish do, however, coincide in the mostly obligative uses of *haber*, while in Gallego and Astur-Leonés they are in most cases of a temporal (prospective) nature, although on a number of occasions – also in Spanish and Catalan – it is difficult to define exactly what is merely futural and what is modal.

Hence, in order to measure these structural differences and their potential incidence on the uses of the periphrases, we conducted a variationist study based on three independent samples made up of texts written in three dialectal areas (monolingual zones, the Catalan-speaking area and the north-western linguistic territories). This material was drawn from a corpus composed of texts that are close to the pole of communicative immediacy. The results of this comparatist study show that the influence of language contact upon the variable under analysis is more complex than a simple descriptive analysis could lead us to believe. Thus, the empirical data seem to confirm the existence of a certain degree of linguistic convergence in the expressive habits of the speakers from the bilingual communities. As is well known, the notion of convergence appears in contact linguistics to account for a diachronic process by which two languages submitted to intense contact reduce the structural distance between them in a certain paradigm of the grammar (Muysken, 1997). As a result of this process, the bilingual speakers of a community display a linguistic behaviour that is different to the corresponding monolingual varieties, although other structural or functional outcomes are also possible (Author, 2006; Muysken, 2000; Sánchez, 2003; Bullock and Toribio, 2004). Despite the existence of an important conceptual heterogeneity in the interpretation of this theoretical notion, this study starts out by considering convergence as “the enhancement of inherent structural similarities found between two language systems” (Bullock and Toribio, 2004: 91) that a particular variety of a language finds itself heading towards as a consequence of the intense contact with another language. Nevertheless, as stated by Poplack et al. (2012: 205), to confirm the influence of contact on a language change phenomenon, it is not enough to simply verify the existence of these diverging usages. It is also necessary to empirically prove that: a) it is, indeed, a

change; b) it was not present in the pre-contact variety; c) it is not present in a contemporaneous non-contact variety; d) it behaves in the same way as its putative borrowed counterpart in the source variety; and e) it differs in non-trivial ways from superficially similar constructions in the host language. How these conditions apply to the case analyzed in these pages?

The findings of this study bear out the complexity that we referred to above. It is true that the data confirm some of these extremes in the expression of verbal modality, although this is not the case in the (far more occasional) sphere of futurity, where the three dialectal groups behave in a practically identical manner. Thus, a feasible explanation for the significantly more extensive use of *haber de* + infinitive among Catalan-speakers could lie in linguistic economy, which might well lead many speakers of Catalan to employ the periphrasis that is common to the two languages used in the community, Spanish and Catalan. In this latter case, moreover, it is the only periphrastic form available. And this same characterisation of convergence as a bilingual optimisation strategy (Muysken, 2002) derives, in the opposite direction, from the reduced number of uses of this periphrasis observed in the north-western bilingual areas. If the most idiosyncratic uses of *haber de* correspond to the domain of futurity in Gallego or Astur-Leonés, it seems logical to expect that, to express modality, which appears far more frequently in the corpus, speakers resort to another periphrasis, *tener que*, which is more common in Spanish and with which it shares similar semantic shades of meaning.

It must be stressed, however, that these differences are mainly limited to the 20th century, because in the previous century the scenario was very similar in the three regions. This also has important theoretical consequences, since it would reveal that the influence of language contact is activated above all in the more advanced stages of language change, that is, those in which one of the variants is already displaying special signs of weakness, as occurs in this case with *haber de*. While this has been the dominant periphrasis in peninsular Spanish for centuries, with similar dialectal usages in different regions, these begin to present distinct behaviours in the moment when *haber de* suddenly drops out of the competition in favour of its rival (*tener que*) and takes refuge within the formal registers of the written language. In that moment, the convergence of languages appears to run in two directions:

a) by intensifying the erosion of the declining variant, whose usages differ in the two languages in contact, or

b) by preserving it to a greater extent than in the other regions, given the formal and semantic similarity among the periphrases in the two grammatical systems and the absence of an alternative variant.

The first seems to be the outcome in the north-western bilingual regions, whereas the conservative influence of the second would characterise the Catalan-speaking areas. Furthermore, this hypothesis would be supported – at least as far as these latter areas are concerned – by the discovery of similar convergence processes in other spheres of grammar. Thus, elsewhere we have seen how the significantly more frequent usages in these Catalan-speaking regions of variants such as [-d-] in words ending in *-ado* (*cantado* vs. *cantao*) or the morphological future (*cantaré* vs. *voy a cantar*), which are submitted to an intense erosion in other dialectal domains of Spanish, could also find their justification in the protective influence of language contact (Author, 2007).

Nevertheless, a deeper analysis shows that those differences, although significant and feasibly explained by the influence of language contact, do not essentially affect the grammar underlying the process of variation and language change

described above. In fact, a systematic study of the variable context surrounding the periphrases shows how the factors constraining this process are substantially similar in all the regions. This is what happens, for example, with the manner of verbal action, the modal values, the length of the verb group and the verb tense and mood, all of which are factors selected as significant by the multivariate analysis in all the dialectal areas that were examined. Even in other cases, such the degree of semantic (im)personalisation, the type of clause, the differences in the stylistic axis, the socio-cultural level or the age, although the support of statistical significance is lost in some groups (an outcome that could perhaps be due to some of the problems of representation of the sample), its frequencies nearly always show identical distributional profiles.

Hence, and bearing in mind these coincidences that point towards an essentially common underlying grammar, the empirical analysis also reveals some differences worthy of mention. Thus, despite the fact that the explanatory ranges are, generally speaking, quite similar among the three dialectal groups, significant exceptions occasionally come to light. This is the case, for instance, of the differences on the stylistic axis, whose explanatory power in the Catalan group is practically three times that of the monolingual areas. In the opposite sense, the time factor obtains a substantially higher range in these latter regions, which reveals, as seen earlier, a greater drop in the usage of the periphrases with *haber* in the period covering the turn from the 19th to the 20th century. Furthermore, some partial differences also reach the probabilistic weight of certain particular factors, where again the singular behaviour of the Catalan group stands out for its greater sense of conservation of the receding variant that was discussed earlier. This is, for example, the case with the tendency to favour *haber de* among the stative verbs and the future indicative, two of the linguistic contexts that traditionally most favour this periphrasis, whose meaning disappears, however, in the other dialectal areas.

Back to the conditions set by Poplack to support the existence of a contact induced change, we must begin to recognize, first, that a) we are indeed facing an old-age change in progress as a consequence of which the periphrasis *tener que* has been replacing *haber de* in modal obligative contexts, and that b) this change has been sharply accelerate during the last century, when an acutely decrease in the frequency of use of *haber de* is observed, both in general terms and practically in all the linguistic and extralinguistic contexts analyzed. Now, we are in a position to confirm that this change takes place in all the Spanish peninsular varieties, including both the monolingual regions and the contact ones. But at the same time, in these former varieties we found some traces that would indicate that the path of variation and change is produced, at least partially, in the same way as its putative counterpart in the source language. Thus, we have seen how the use of *haber de* is maintained significantly more in the linguistic Catalan areas, in which the autochthonous language lack a parallel counterpart. At the same time, this bilingual community shows some singularities in the explanatory power of the stylistic axis, as well as in the factor weight of some isolated structural constraint. Additionally, just the opposite outcome is found among the North-west regions, where languages such as Gallego or Astur-Leonés limit the canonical uses of this periphrasis to temporal non modal meanings, which could explain its bigger erosion in the Spanish of these areas.

Nevertheless, bearing all these facts in mind, we have also seen how the potential explanation of linguistic contact as an inducing factor in this case of change in progress is seriously moderated by other no less important findings. Thus, in the temporal axis, we have seen how the differences observed only appear in the most advanced stage of change, when the erosion of the receding variant has (apparently)

reached the point of no return. But even in this case, we have found that, with the odd exception, the envelop of variation is basically the same in all areas.

In sum, the implementation of the principles and methods of the comparative variationist framework on a text corpus lying close to the pole of communicative immediacy has enabled us to confirm the influence of language contact on the process of language change that has affected the modal periphrasis *haber de* + infinitive throughout the last two centuries. This influence, however, can be considered moderate and has a scarce effect on the grammar underlying this phenomenon of variation and change in the different dialectal areas analysed, which essentially behave in a similar way.

References

- Academia de la Llingua Asturiana. (2001). *Gramática de la Llingua Asturiana*. Uviéu, 3ª edición.
- Álvarez Caccamo, Celso. (1983). A influencia do gallego sobre o sistema verbal e no uso de certas perífrases verbais do castelán de Galiza, *Grial*, 82. 423-442.
- Álvarez, Rosario, Xosé Luís Regueira & Henrique Monteagudo. (1998). *Gramática Galega*. Vigo: Galaxia. 7ª edición.
- Autor 2004. _____
- Autor 2005 _____
- Autor 2006 _____
- Autor 2008 a _____
- Autor 2008 b _____
- Autor (in press).
- Autor y Autor X a (in press) _____
- Autor y Autor X b (in press) _____
- Autor y Autor Y (in press) _____
- Badia, Antoni M. 1983. *Gramática catalana*. Madrid: Gredos.
- Balasch, Sonia. (2008). *Debe (de) ser*: evolución de la variación. In Maurice Westmoreland & Juan Antonio Thomas (eds.), *Selected Proceedings of the 4th Workshop on Spanish Sociolinguistics*. Somerville MA: Cascadilla Proceedings Project. 109-119.
- Balasch, Sonia. (2012). Revisión amplia de la variación *deber(de)*+Infinitivo en la diacronía del español. In Emilio Montero (ed.), *Actas del VIII Congreso Internacional de Historia de la Lengua Española*, Santiago de Compostela, España. Septiembre, 2009. 607-620.
- Bergs, A. 2012. The Uniformitarian Principle and the Risk of Anachronisms in Language and Social History. In Juan Manuel Hernández-Campoy & Juan Carlos Conde Silvestre (eds.) *The Handbook of historical sociolinguistics*. Malden, MA. John Wiley & Sons Inc. 80-98.
- Chambers, Jack. K. (1995). *Sociolinguistic Theory*. Oxford: Blackwell.
- Conde Silvestre, Juan C. (2007). *Sociolingüística histórica*. Madrid: Gredos.
- Davis, Mark. (2002). Un corpus anotado de 100.000.000 palabras del español histórico y moderno. In *Procesamiento del lenguaje natural*, 29: 21-27
- Davis, Mark. (2006). *A frequency dictionary of Spanish*. New York: Routledge.
- De Bruyne, Jacques. (1993). *Spanische Grammatik*. Tübingen: Niemeyer.
- Elpass, Stephan. (2012). The use of private letters and diaries in sociolinguistic investigation. In Juan Manuel Hernández-Campoy & Juan Carlos Conde Silvestre (eds.), *The Handbook of historical sociolinguistics*. Malden, MA : John Wiley & Sons Inc. 156-169.
- Fernández de Castro, Félix. (1999). *Las perífrasis verbales en el español actual*. Madrid: Gredos.
- Garachana Camarero, Mar. (1997). Acerca de los condicionamientos cognitivos y lingüísticos de la sustitución de 'aver' por 'tener'. *Anuario Galego de Filoloxia*, 24. 203-235.
- Garachana Camarero, Mar & Malte Rosenmeyer. (2011). Rutinas léxicas en el cambio gramatical. El caso de las perífrasis deónticas e iterativas. *Revista de Historia de la Lengua Española*, 6. 35-60.
- García, Constantino (1976). Interferencias lingüísticas entre gallego y castellano. *Revista Española de Lingüística* 6:2. 327-343.
- García Fernández, Luis. (2006). *Diccionario de perífrasis verbales*. Madrid: Gredos.
- García Fernández, Luis. (2012). *Las perífrasis verbales en español*, Madrid: Castalia.
- Gili Gaya, Samuel. (1970). *Curso superior de sintaxis española*. Barcelona: Vox (1ª edición: 1961).

- Gómez Torrego, Leonardo. (1988). *Las perífrasis verbales*. Madrid: Arco Libros.
- Gómez Torrego, Leonardo. (1999). Los verbos auxiliares. Las perífrasis verbales de infinitivo. In I. Bosque & V. Demonte (dirs.), *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa-Calpe. 3223-3390.
- Gutiérrez Ordóñez, Salvador (1980). Tengo que vender unos libros / Tengo unos libros que vender. *Verba*, 7. 389-396.
- Hernández Díaz, Axel. (2006). Posesión y existencia. La competencia de haber y tener y haber existencial. In Company Concepción (coord.), *Sintaxis histórica de la lengua española*, Vol. 1, Tomo II (Primera parte, la frase verbal). 1055-1164.
- Hernández García, Carmen. (1998). Una propuesta de clasificación de la interferencia lingüística a partir de dos lenguas en contacto: el catalán y el español. *Hesperia*, 1. 61-80.
- Iglesias Feijó, Encarna. (1969). *La futuridad en Celta*. Santiago de Compostela: Universidade de Santiago de Compostela.
- Kabatek, Johannes. (1996). *Die Sprecher als Linguisten. Interferenz- und Sprachwandelphänomene dargestellt an alicischen der Gegenwart*. Tübingen: Niemeyer.
- Kany, Charles E. (1969). *Sintaxis hispanoamericana*. Madrid: Gredos.
- Keniston, Hayward. (1936). Verbal Aspect in Spanish. *Hispania*, 19. 163-176.
- Labov, William (1972). *Sociolinguistics Patterns*. Philadelphia: Pennsylvania U.P.
- Labov, William. 2001. *Principles of Linguistic Change: External Factors*. Philadelphia: Blackwell Publishing Co.
- Lapesa, Rafael. (2000). *Estudios de morfosintaxis histórica del español*. Edited by Rafael Cano & M^a Teresa Echenique, Madrid: Gredos.
- López Izquierdo, Marta. (2008). Las perífrasis modales de necesidad: emergencia y renovación. In C. Company & J. G. Moreno (coords.), *Actas del VII Congreso Internacional de Historia de la Lengua española. Yucatán, 4-8 de septiembre de 2006*, Vol. 1, 2008. 789-806.
- Martínez Díaz, Eva. (2002). Las perífrasis modales de obligación "tener que + infinitivo" y "haber de + infinitivo" : Variación e interferencia en el español de Barcelona. Ph. D. dissertation. Available at: http://www.tdx.cat/bitstream/handle/10803/1713/01.EMD_1de6.pdf;jsessionid=249E33DE1A739FC654A04305576ECD1F.tdx2?sequence=1
- Martínez Díaz, Eva. (2003). La frecuencia de uso de haber y tener en las estructuras perifrásticas de obligación. Algún fenómeno de variación en el español de Cataluña. *Interlingüística*, 14. 681-694.
- Martínez Díaz, Eva. (2008). La expresión de la modalidad de obligación en el Corpus del español conversacional de Barcelona y su área metropolitana (2001). In A. Moreno (coord.), *El valor de la diversidad (meta)lingüística. Actas del VIII congreso de Lingüística General*. Available at: <http://www.lllf.uam.es/clg8/actas/index.html>.
- Marvà, Jeroni. (1983). *Curs superior de gramàtica catalana*. Barcelona: Barcino.
- Matsuda, K. (1993). Dissecting analogical leveling quantitatively: The case of the innovative potential suffix in Tokyo Japanese. *Language in Society*, 18. 213-234.
- Muysken, Peter. (2000). *Bilingual Speech*. Cambridge: Cambridge University Press.
- Muysken, Peter. (2002). *Modelling bilingual speech: Otra vez Ms.*, Radboud: University Nijmegen.
- Nevalainen, Tertu. & Helena Raumolin-Brnberg. (2003). *Historical Sociolinguistics: Language Change in Tudor and Stuart England*. London: Longman Pearson Education.
- Nevalainen, Tertu. & Helena Raumolin-Brnberg. (2012). Historical sociolinguistics: origins, motivations, and paradigms. In Juan Manuel Hernández-Campoy & Juan Carlos Conde Silvestre (eds.) *The Handbook of historical sociolinguistics*. Malden, MA. John Wiley & Sons Inc.. 22-40.
- Olbertz, Hella (1998). *Verbal periphrases in a Functional Grammar of Spanish*, Berlin: Mouton de Gruyter.
- Oesterreicher, Wulf. (2004). Textos entre immediatez y distancia comunicativas. El problema de lo hablado escrito en el Siglo de Oro. In Rafael Cano Aguilar (coord.), *Historia de la lengua española*, Barcelona: Ariel. 729-769.
- Payrató, Lluís. (1985). *La interferència lingüística (Comentaris i exemples català castellà)*. Barcelona: Ed. Curial-Publicacions de l'Abadia de Montserrat.
- Poplack, Shana. & Marjory Meechan. (1998). Introduction: How languages fit together in codemixing. Instant loans, easy conditions: The productivity of bilingual borrowing. *Journal of Bilingualism*, 2 (2). 127-138.
- Poplack, Shana, Lauren Zentz & Nathalie Dion. (2012). Phrase-final preposition in Quebec French: An empirical study of contact, code-switching and resistance to convergence. *Bilingualism: Language and Cognition* 15, 2. 203-225.
- Pountain, Christopher J. (2001). *A History of the Spanish Language through Texts*. New York: Routledge.
- Roca Pons, José. (1980). *Introducción a la gramática*, Madrid: Vergara.

- Rojo, Guillermo. (1974). *Las perífrasis verbales en el gallego actual*. Anejo 2 de la revista *Verba*.
- Rojo, Guillermo. (2004). El español de Galicia, In Rafael Cano Aguilar (coord.), *Historia de la lengua española*, Barcelona: Ariel. 1087-1101.
- Sáez-Godoy, Leopoldo. (1968). Algunas observaciones sobre la expresión del futuro en español, In *Actas del XI Congreso Internacional de Lingüística y Filología Románicas IV*. Madrid, C.S.I.C. 1875-1889.
- Seco, Manuel. (1986). *Diccionario de dudas y dificultades de la lengua española*. Madrid: Espasa-Calpe.
- Seifert, Eva. (1930). *Haber y tener* como expresiones de la posesión en español. *RFE*, XVII. 233-276 y 345-385.
- Sinner, Carsten. (2003) Valores y usos de haber de + infinitivo y tener que + infinitivo en diferentes variedades del castellano. In Pusch, Claus & Andreas Wesch (eds.), *Perâifrasis verbals en les llengües (ibero-)romániques*, Madrid: Iberoamericana-Vervuert. 193-206.
- Sinner, Carsten. (2004). *El castellano de Cataluña*, Tübingen: Max Niemeyer Verlag.
- Sinner, Carsten & Andreas Wesch (eds.). (2008). *El castellano en las tierras de habla catalana*. Frankfurt and Madrid: Vervuert and Iberoamericana.
- Steel, Brian. (1982). Algunos apuntes para un nuevo diccionario de americanismos. In M. Perl (ed.), *Estudios sobre el léxico del español en América*, Leipzig: Enzyklopädie. 176-197.
- Stengaard, Birte. (2003). Función y evolución de la construcción aver + preposición + infinitivo. In J. L. Girón & J. J. de Bustos (coords.), *Actas del VI Congreso Internacional de Historia de la Lengua Española. Madrid 29 de septiembre-3 de octubre de 2003*, Vol. 1, 2006. 1147-1152.
- Tagliamonte, Sali. (2012). *Variationist Sociolinguistics: Change, Observation, Interpretation*. Oxford: Wiley-Blackwell Publishers.
- Tarallo, F. (1989). Inside and outside relative clauses. In R. Fasold & D. Schiffrin (eds.), *Language change and variation*. Washington: Georgetown University Press. 255-274.
- Vann, Robert. (2002). El castellà catalanitzat a Barcelona: Perspectives lingüístiques i culturals. *Catalan Review*, 15, 1. 117-131.
- Wesch, Andreas. (1994). Bereicherung und Nivellierung semantischer Strukturen durch Interferenzen am Beispiel Spanisch / Katalanisch. In Helfrich, Uta & , Claudia Maria Riehl (eds.). *Mehrsprachigkeit in Europa - Hindernis oder Chance?* Wilhelmsfeld: Egert. 165-178
- Wesch, Andreas (1997). El castellano hablado en Barcelona y el influjo del catalán. Esbozo de un programa de investigación. *Verba. Anuario Galego de Filoloxia*, 24. 287-312.
- Westmoreland, Maurice. (1997). The dialectalization of Spanish future tense usage. *Word* 48:3. 375-395.
- Yllera, Alicia. (1980). *Sintaxis histórica del verbo español: Las perífrasis medievales*. Zaragoza: Universidad de Zaragoza.